

Plano de análise:

EDA dos dados de covid 19 do estado de alagoas

Autor: Alcides Coutinho

Contato:

Objetivo: Explorar o comportamento da covid 19 no estado de Alagoas com os dados disponibilizados pelo governo.

Análise no estado, na capital e no interior

1. Casos e óbitos:
 - a. Quantos casos e óbitos / porcentagem disso?
 - b. Qual o gênero?
 - c. Qual a faixa etária atingida nos casos e nos óbitos?
2. Municípios
 - a. Quais municípios tiveram mais mortes?
 - b. Qual a taxa de mortalidade a cada 1000 habitantes desses municípios?
 - c. Qual a faixa etária dos casos desses municípios e onde se concentra?

Análise das Situações Confirmadas

1. Distribuição
 - a. O que podemos interpretar da distribuição da faixa etária das situações?
 - i. Outliers?
 - ii. Concentração da faixa etária?
 - iii. Outras informações importantes?

Análise dos óbitos

1. Comorbidades
 - a. Porcentagem em relação total?
 - b. Quais as principais? Quanto é a porcentagem da presença?
 - c. Faixa etária dos que não tiveram comorbidades e sua concentração?

Análise Temporal

1. Comportamento
 - a. Comportamento dos casos e das mortes ao longo dos anos?
2. Series temporais
 - a. Qual é a atual tendência?
 - b. Há sazonalidade?
3. O comportamento dos casos e óbitos a partir da 1 e 2 doses das vacinas
 - a. Houve queda?
 - b. Qual era a média de mortes e casos antes e depois das doses?

Introdução

O vírus covid 19 causou mais de 6 milhões de mortes e diversos outros danos, sendo eles psicológicos e econômicos alterando a forma como a humanidade se comporta e pensa. A fim de entender melhor o comportamento desse vírus, utilizei os dados do meu estado, Alagoas, para poder fazer uma análise exploratória dos dados com base no plano de análise, mostrado anteriormente.

Uma rápida introdução sobre alagoas: Localizada no nordeste do Brasil é o 25º maior estado, tendo uma população estimada em 3,351 milhões de habitantes com 102 municípios, tendo como a capital Maceió com uma população maior que 1 milhão de habitantes.

Todo trabalho de ciência dos dados é feito com base em alguma pergunta, sendo assim, o plano de análise levanta quais as perguntas feitas que buscamos responder com os dados disponíveis.

Esse Artigo vai tratar de demonstrar somente a análise feita com base nos dados, toda a parte de limpeza dos dados e códigos estão disponíveis no github, onde esse artigo também se encontra em pdf.

Dados

Os dados disponibilizados pelo governo do estado de alagoas, demonstra que ocorreram 300.044 casos de covid 19, com data de registro de 26/03/2020 até 12/06/2022. Os dados são compostos das seguintes *features*:

- ❖ - **Idade** – Idade de cada paciente
- ❖ - **Sexo** – Sexo de cada paciente
- ❖ - **Cidade** – Qual a cidade alagoana
- ❖ - **Comorbidade** – Quais comorbidades o paciente possui
- ❖ - **Situação do paciente** – Aqui informa se o paciente foi recuperado / isolamento social / óbito.
- ❖ - **Classificação** – Só foi disponibilizado os dados dos pacientes confirmados com covid.
- ❖ - **Data de serviço** – O dia do atendimento
- ❖ - **Data de óbito** – Data do óbito dos pacientes de faleceram
- ❖ - **Data de confirmação do exame** – Data do resultado positivo

Toda a parte de limpeza dos dados, transformações e tratamentos será tratado em outro artigo, também presente no github.

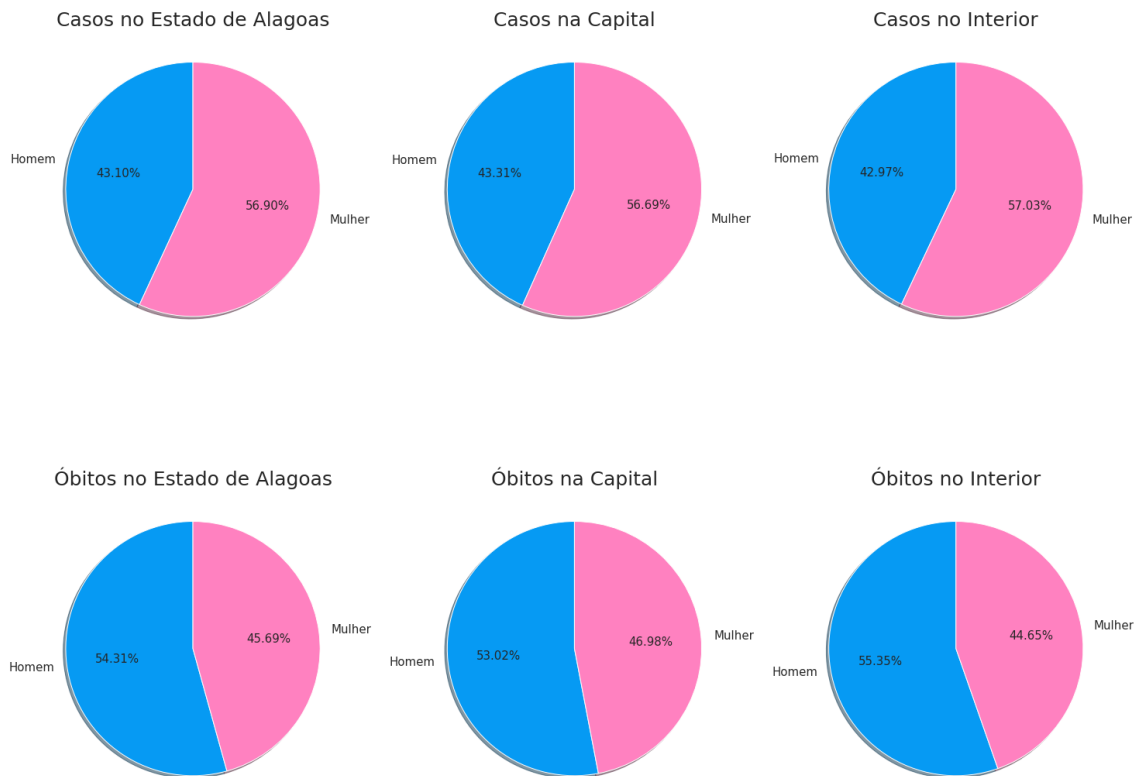
Análise no estado, na capital e no interior

Casos e óbitos:



Com a imagem acima podemos verificar algumas coisas, sendo ela o número de casos, tanto no total quanto na capital e no interior, e pode-se verificar a quantidade de óbitos, tanto em valores absolutos quanto a proporção dos óbitos em relação aos casos em alagoas.

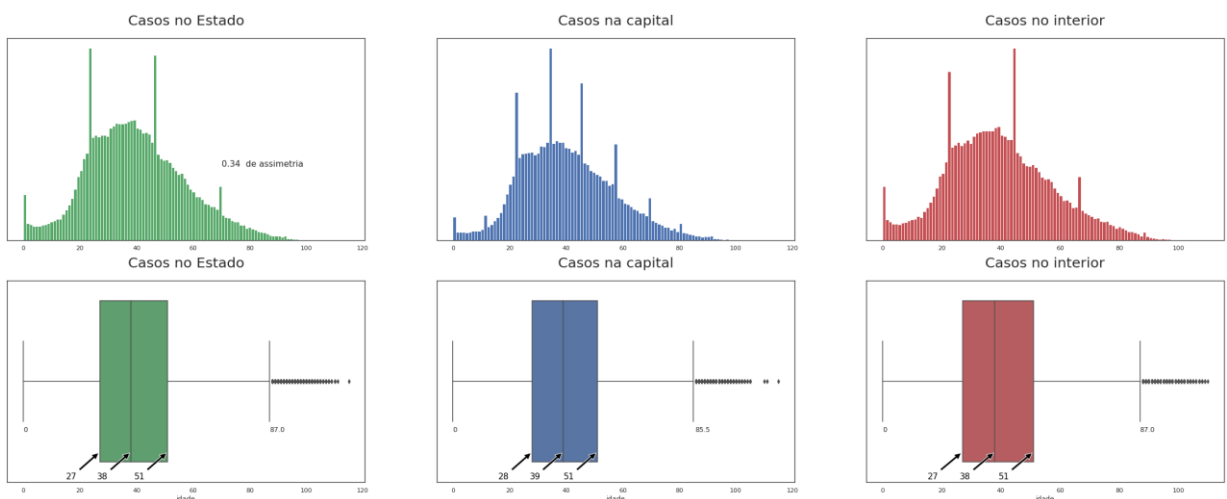
- ❖ O gráfico demonstra que a capital teve uma maior proporção de óbitos x casos do que o interior
- ❖ O interior apresentou mais mortes e casos, mas com uma proporção óbitos casos menor.



Na imagem acima foi feita a divisão com base no sexo dos pacientes, analisando as imagens é possível chegar as seguintes conclusões:

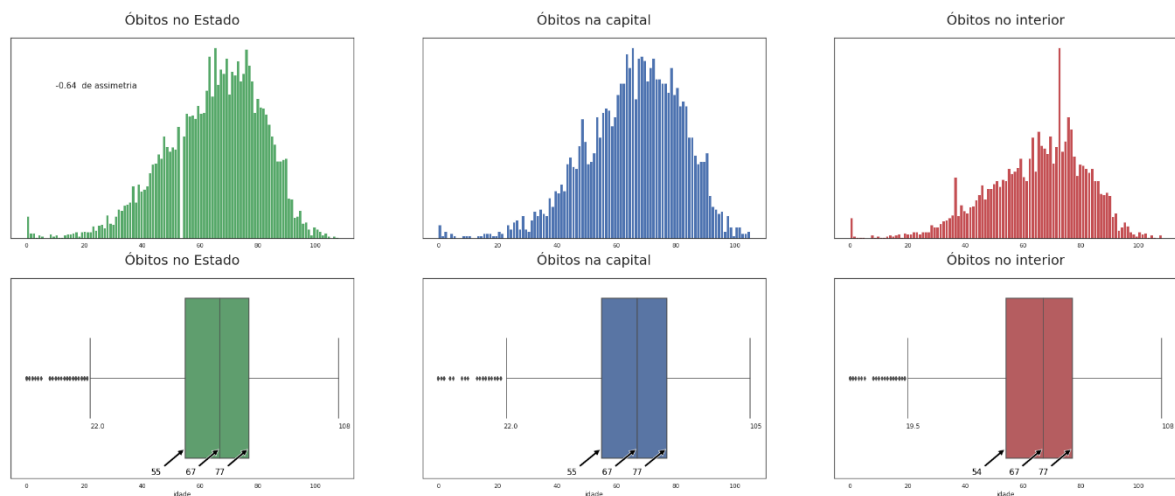
- ❖ Mais mulheres foram contaminadas do que os homens
- ❖ Mais homens faleceram, mesmo apresentando menores casos.
- ❖ As ocorrências foram bem semelhantes tanto na capital quanto no interior

Algumas hipóteses podem ser levantadas para esses casos, como o fato das mulheres terem a tendência de procurar auxílio médico mais comumente do que os homens, mas não há como extrair isso dos dados, então só posso ficar no campo da especulação.



Analisando primeiro a distribuição das idades das pessoas contaminadas, temos que os números foram bem idênticos na capital e no interior, tendo como concentração maior de contaminados na faixa dos 27 aos 51 anos. Outra informação retirada do gráfico é o valor da assimetria, que na estatística inferencial demonstra que há uma certa tendência dos dados de estarem inclinados, seja para esquerda ou direita, com base nisso têm-se o valor assimétrico para a direita, concluindo então que as pessoas que pegaram a covid 19 foram as mais jovens. Outra análise interessante é pelo

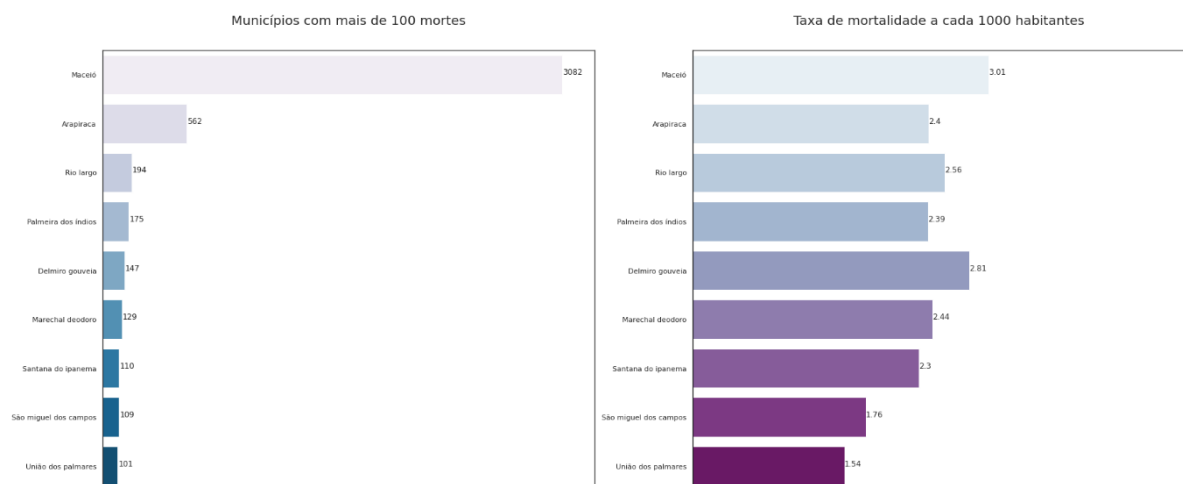
outliers, ou pontos fora da curva, que são os dados que ocorrem de forma incomum e aqui temos pessoas acima dos 87 anos, talvez devido as precauções tomadas e a quantidade de pessoas nessa idade tenha sido calculado como outliers.



Agora avaliando a distribuição da idade nos óbitos da covid 19 observa-se uma clara inclinação a direita, ou seja, conclui-se que quanto maior é a idade, maior será o número de óbitos. A concentração dos óbitos ficou na faixa de 55 até 77 anos de idade, e temos um dado muito valioso, que são os outliers. Aqui temos uma clara diferença no interior e na capital, na capital as idades inferiores a 22 anos sofreram menos óbitos, então classificadas como outliers pelo cálculo do interquartil, já no interior o limite inferior foi até 20 anos, ou seja, foi mais comum o falecimento de pessoas de 20 anos para cima, enquanto na capital subiu um pouco mais.

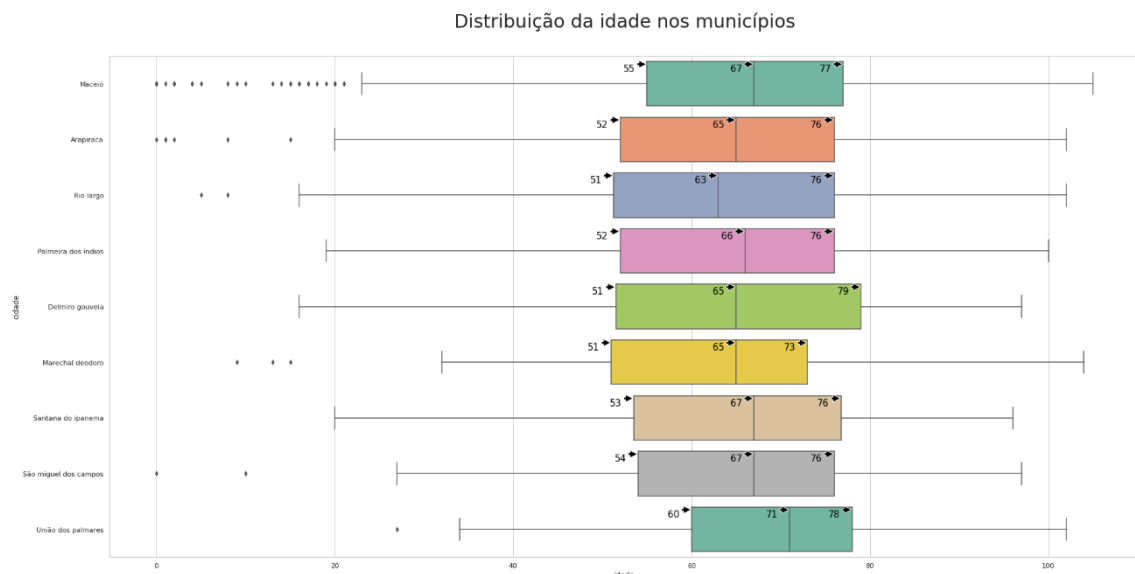
Municípios

Como alagoas apresenta a todo 102 municípios, fazer uma análise desses todos acaba sendo trabalhosa e pouco produtiva, pelo fato de muitos desses municípios terem menos de 100 mil habitantes. Sendo assim, irei selecionar somente os municípios que tiveram mais de 100 óbitos e irei analisar em cima disso.



Como mostrado na imagem acima, esses foram os 9 municípios que registraram mais de 100 óbitos, como era de se esperar, as 2 maiores cidades do estado estão em 1 e 2 em óbitos, como algumas regiões metropolitanas aparecendo, como Marechal Deodoro e Rio Largo.

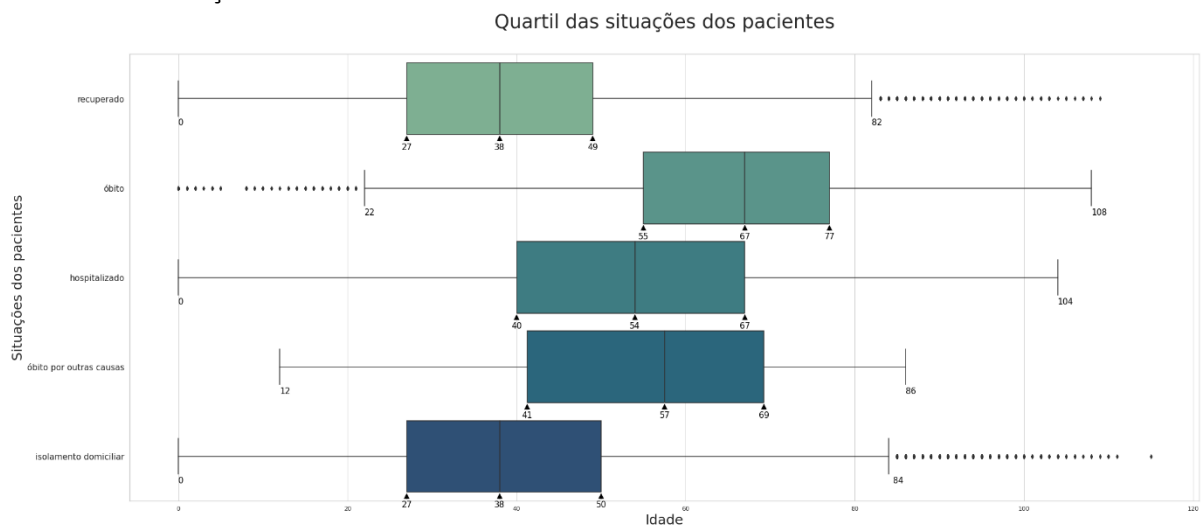
Com base nas mortes podemos calcular o coeficiente e a taxa de mortalidade da covid 19, assim apresentando Maceió como a maior taxa de mortalidade por 1000 habitantes, e como surpresa a cidade de Delmiro Gouveia aparece como a segunda maior taxa, mesmo sendo uma cidade mais ao interior do estado e longe da capital.



Com esses dados de distribuição é possível observar que a partir de 50 anos é que se concentra os óbitos nos municípios e que os outliers se concentram abaixo de 20 anos.

Análise das Situações Confirmadas

Distribuição



Utilizando o modelo de interquartil/quartil para avaliar uma distribuição de dados é uma das melhores formas de visualizar outliers, que são os pontos fora da curva. Esses pontos são considerados incomuns na distribuição desses dados. Outra informação importante fornecida é a caixa, dentro dela está contida 50 % dos dados, e do limite inferior (a primeira barra do lado esquerdo) até o final da caixa, temos 75% dos dados. Sendo assim, podemos avaliar as seguintes informações presentes no gráfico.

- ❖ Temos 5 situações registradas dos pacientes: recuperado, hospitalizado, isolamento domiciliar, óbito por outras causas e óbitos. Como o óbito por outras causas não nos interessa no momento, iremos deixá-la de lado.
- ❖ Recuperados – Podemos observar que pessoas acima de 82 anos foram consideradas fora da curva, ou seja, acima de 82 anos a recuperação não foi registrada, isso se deve provavelmente que esses pacientes foram hospitalizados devido a idade avançada, então e devem ter caído em outra classificação, como hospitalizado, para ter um cuidado mais próximo. Temos que a concentração de pessoas recuperadas ficou na faixa de 27 – 49 anos de idade.
- ❖ Óbito – Aqui temos os pontos fora da curva abaixo de 22 anos, ou seja, pessoas jovens tendem a sobreviver à covid 19, isso sendo uma excelente constatação. Aqui já demonstra que os óbitos ficaram concentrados na faixa de 55-77 anos de idade, demonstrando que atinge pessoas entrando e na terceira idade.
- ❖ Hospitalizados – Todas as idades ficaram dentro do range de hospitalização, como a covid se comporta de diversas formas, então é um dado interessante.

- ❖ Isolamento domiciliar – constatando a leitura dos recuperados, idosos acima de 84 anos enviados para isolamento domiciliar foram bem baixos, reforçando a hipótese levantada anteriormente.

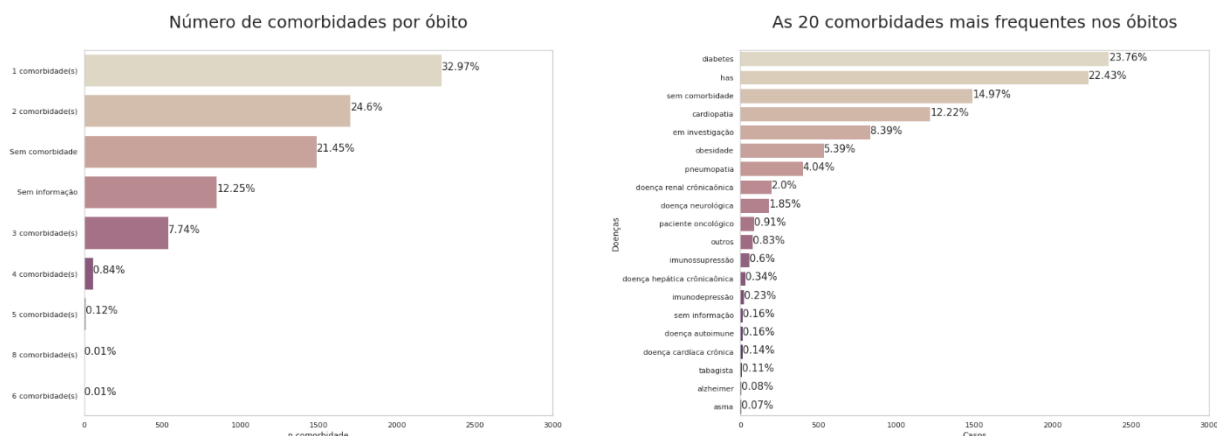
Análise de óbitos

Comorbidades



Tivemos mais de 100 doenças classificadas como comorbidades e para o entendimento da importância da comorbidade em relação aos óbitos, então unifiquei os dados para obter quem tem e não tem comorbidade e deixei claro os dados que não apresentam informações. O que podemos obter com os dados são:

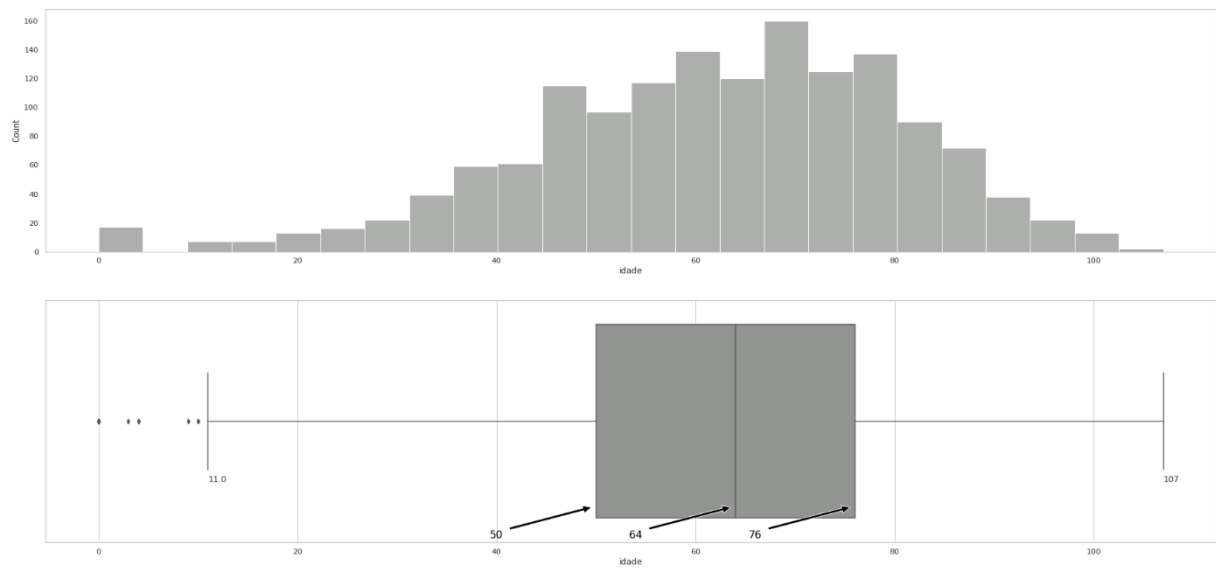
- ❖ Fica evidente olhando para o gráfico do estado a relevância da comorbidade, já que está presente em 66,29% dos óbitos. Se os dados sem informação seguirem a mesma distribuição dos dados com e sem comorbidade, podemos ter em torno de 70% dos casos.
- ❖ Na capital fica evidente a correlação de óbitos com as comorbidades
- ❖ Comparando a capital com o interior, vemos que há uma clara diferença entre pessoas com comorbidades e sem, podemos então levantar algumas hipóteses (na base do achismo) como a falta de estrutura e exames de rotina para descobrir comorbidades e a vida mais saudável do interior longe da capital.
- ❖ A falta de informação é maior do que na capital em uma margem bem considerável, isso pode ser relacionado com as estruturas mais precárias do interior que dificultou obter esses dados no início da pandemia (hipótese)



Como mais de 60% dos óbitos tinham algum tipo de comorbidade, é importante saber quais comorbidades eram mais presentes, então analisando os gráficos podemos compreender que:

- ❖ Analisando o primeiro gráfico, fica evidente que a maioria das pessoas tinha 1 ou 2 comorbidades, isso acaba sendo um sinal de alerta para o perigo desse vírus para pessoas que apresentam algum tipo de comorbidade.
- ❖ Mas para saber o grupo de risco, plotei o gráfico mostrando as 20 mais, me limitei a 20 pelo fato de a ocorrência das outras doenças serem bastante baixa, como já apresentada no gráfico algumas com baixa ocorrência. Sendo assim, podemos tirar que as doenças mais perigosas são:
 - a. Diabetes
 - b. Hipertensão (has)
 - c. Cardiopatia
 - d. Obesidade

Distribuição dos óbitos sem comorbidade



Como os óbitos sem comorbidades foram superiores a 20%, é importante avaliar esses casos com os dados possíveis, sendo uma possível forma pela idade, assim podemos ver a faixa de idade para os óbitos.

- ❖ Óbitos se concentraram para pacientes acima de 50 anos.
- ❖ Abaixo de 11 anos foi considerado outlier.